

# Data Warehousing and Data Mining Capabilities Demonstration



**08-25-04**

**Steve Ohman  
(505) 678-8243**

[ohmansm@trac.wsmr.army.mil](mailto:ohmansm@trac.wsmr.army.mil)

**Regina Berg  
(505) 678-7089**

# Purpose

---

**Prepare data mining capability providers for TRAC-WSMR's expectations and requirements for the upcoming demonstrations.**

## Overall Goal:

**TRAC-WSMR must be able to access and query 500GB by 50 users simultaneously.  
Additionally, the time from initiate complex query to results should be less than three minutes.**

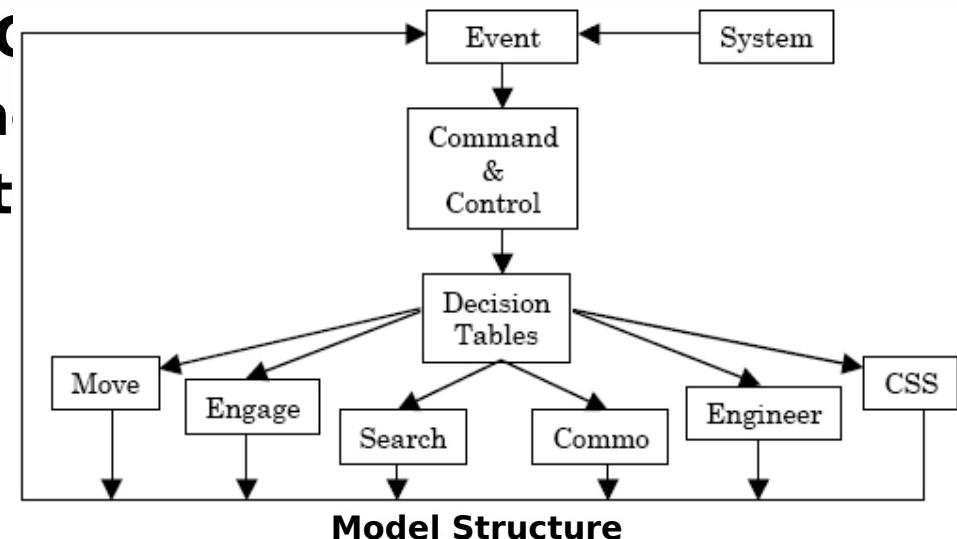
# **Agenda**

---

- **Background**
- **Sample File Types**
- **File Types Relationships**
- **Sample Queries (MoE)**
- **Requirements for Data Mining Tool**
- **Expected Deliverables**

# Background

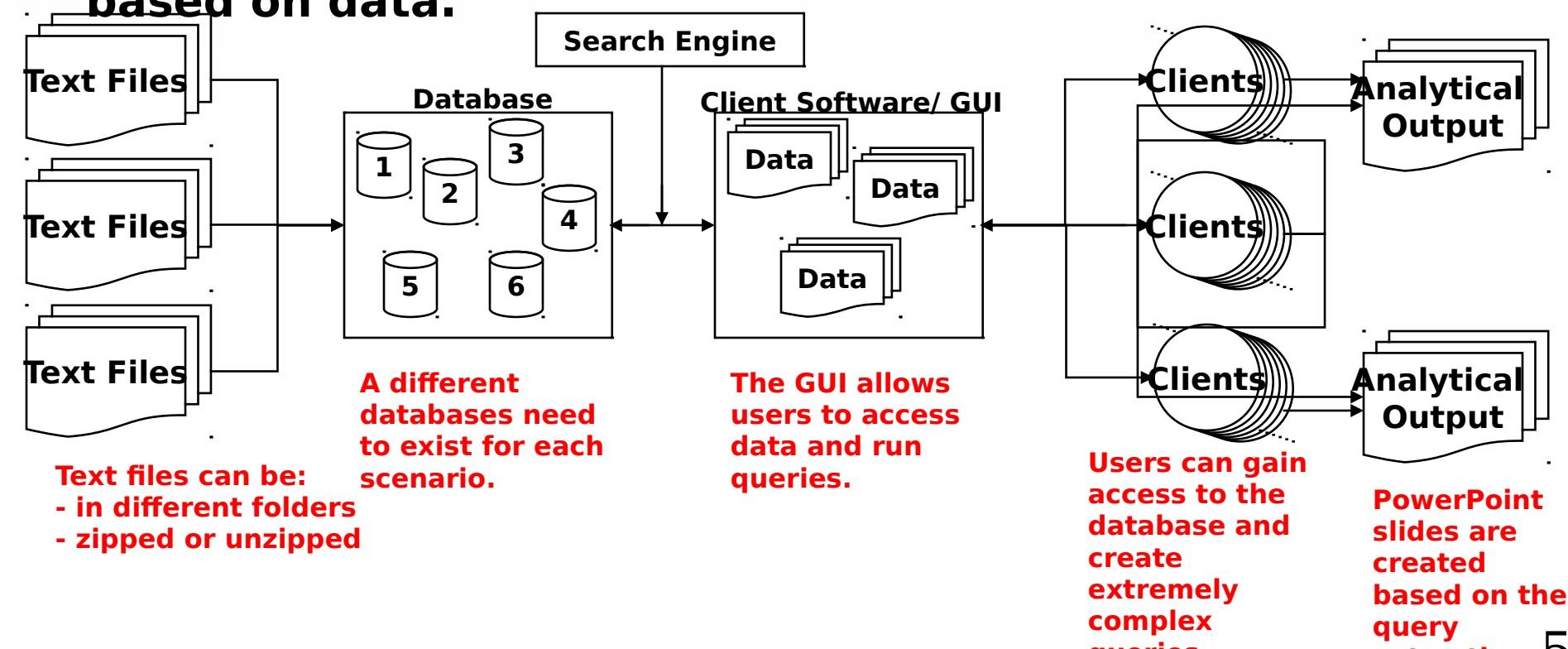
- The text files used are sample model output from the Army's Model and Simulation program, Combined Arms and Support Task Force Evaluation Model (CASTFOREM).
- CASTFOREM is a stochastic model and is comprised of and logs data for:
  - Command and Control (C&C)
  - Communications (Commo)
  - Combat Service Support (CSS)
  - Engineering
  - Surveillance (Search)
  - Engagements
  - Maneuvers
  - System/Environment
- CASTFOREM is highly flexible both as to what it can model and the resolution an object or process contains.



# Background

## Basic database and data mining envision:

- Model outputs various text files.
- Imports text files into database
- Users access the database through a GUI to extract data.
- Users generate and format reports, charts, and slides based on data.



# **Sample File Types (1 of 11)**

---

**TRAC WSMR has several model which generate various data various formats. There are well over 100 different tables which need to be included in analysis. The following are sample output files a model can generate:**

## **Sample output files**

- Engagement Summary (ES)**
- Mission Thread (MT)**
- Common Operating Picture (COP)**
- Unit List (UL)**
- Unit Summary (US)**
- Tactical Area (TA)**

# Sample File Types (2 of 11)

**Engagement Summary (ES) file contains information for all forces engagements (shots, hits, kills) and acquisitions.**

DATATYPE	Acquisition or Engagement
REP	Replication Number
FIRETIME	Firing Event Time
TIME	Detection/Impact Event time
FSIDE	Firer/Detector Side
FUNIT	Firing Unit ID Number
TFUNIT	Firing Unit Type
FIREX	Firing Unit X Coordinate
FIREY	Firing Unit Y Coordinate
FIREZ	Firing Unit Z Coordinate*
SENSOR	Firing Unit Sensor Type
MUNITION	Firing Unit Munition Type
TSIDE	Target Side
TUNIT	Target Unit ID Number
TTUNIT	Target Unit Type
TGTX	Target Unit X Coordinate
TGTY	Target Unit Y Coordinate
TGTZ	Target Unit Z Coordinate*
FSPD	Firing Unit Speed
TSPD	Target Unit Speed

<u>Field</u>	<u>Description</u>
FIRRNG	Firing Unit Range
DESRNG	Designator Range
ACQ	Acquisition Level
PHIT	Probability of Hit
PKILL	Probability of Kill Given Hit
MISSDIST	Impact Distance From Target (IF)
IMPACTX	Impact X Coordinate
IMPACTY	Impact Y Coordinate
ASPECT	Aspect Angle to Firer FROM TARGET
HM	Hit or Miss Type
KILL	Kill Type
PRNKILL	Quantity of Personnel Killed
RPB	Rounds Per Burst
RPE	Rounds Per Engagement
TCUNIT	Carrier Unit Type (Mounted Personnel)
CUNIT	Carrier Unit ID (Mounted Personnel)
PI	Probability of Incapacitation
SHF	Sheaf Type

# Sample File Types (3 of 11)

**Mission Thread (MT) file contains data on the known life of each target from when it was first acquired to when it was affected.**

<u>Field</u>	<u>Description</u>	<u>Field</u>	<u>Description</u>
REP	Replication run number	TRACK_UNIT	Bumper number of unit
GAME_TIME	Game time of event	TRACK_TYPE	Unit type of unit
SENDER_NAME	Bumper number of sender	TRACK_X	Units perceived x coordinate
SENDER_TYPE	Unit type of sender	TRACK_Y	Units perceived y coordinate
SENDER_X	x coordinate sender	MUNITION_NAME	munition name
SENDER_Y	y coordinate sender	MSN_CODE	Code number of Artillery
DESIGNATOR_NAME	Bumper number of unit designating target	mission	
DESIGNATOR_TYPE	Unit type of the designator	GEN_TIME	Time Artillery mission is generated.
DESIGNATOR_X	x coordinate of the	FO_NAME	Bumper number of Forward Observer
designator		FO_TYPE	Unit type of Forward Observer
DESIGNATOR_Y	y coordinate of the	TARGET_X	Ground truth x coordinate of
designator		TARGET_Y	Ground truth y coordinate of target
RECEIVER_NAME	Bumper number of receiver	AIM_X	x coordinate of Artillery aim
RECIEVER_TYPE	Unit type of receiver	AIM_Y	y coordinate of Artillery aim
RECEIVER_X	x coordinate of message	IMPACT_X	x coordinate of Artillery impact
receiver		IMPACT_Y	y coordinate of Artillery impact
RECEIVER_Y	y coordinate message	ACQ_Level	Acquisition level of Target
receiver		Kill_Level	Perceived or actual Kill level
TRACK_ID	Number given to unit being tracked.		

# Sample File Types (4 of 11)

---

**Common Operation Picture (COP) file contains perceived situational data from a blue force perspective.**

<u>Field</u>	<u>Description</u>
REP	Replication Number
COP_TIME	COP push time
MSG_TIME	Fusion time or FSA update time
REC_TIME time	Track initiation time, mission generation time, or kill assessment
COP_UNIT	COP Receiver (e.g. BCP10000, 11000, 12000, 13000, BC14110)
SIDE	Report of FRIEND or THREAT
UNIT_ID	bumper number of reported unit
TYPE_TUNIT	target unit type
ACT_X	Actual x grid
ACT_Y	Actual y grid
PER_X	Perceived x grid
PER_Y	Perceived y grid
DIST_ERROR	Perceived - Actual distance
REP_UNIT	bumper number of reporting unit
TYPE_RUNIT	reporting unit type
BDA	Battle Damage Assessment
ACQ_LEVEL	Best Acquisition
TRACK_ID	ID number given to each unit being tracked (Identical to Mission

# Sample File Types (5 of 11)

---

**Unit List (UL) file contains a listing of all units in the scenario.**

<u>Field</u>	<u>Description</u>
<b>Side</b>	<b>Side of the unit</b>
<b>Unit_ID</b>	<b>Bumper number of unit</b>
<b>Type_Unit</b>	<b>System Type of Unit</b>
<b>X</b>	<b>Starting X coordinate of unit</b>
<b>Y</b>	<b>Starting Y coordinate of unit</b>

# Sample File Types (6 of 11)

---

**Unit Summary (US) file contains a listing of all unit types with their sensors, munitions, and radios.**

<u>Field</u>	<u>Description</u>	<u>Field</u>	<u>Description</u>
<b>Side</b>	<b>Side of the unit</b>	<b>Inf_Name</b>	<b>not used</b>
<b>Profile</b>	<b>not used</b>	<b>Inf_Code</b>	<b>not used</b>
<b>Frequency</b>	<b>not used</b>	<b># Inf_Sqds</b>	<b>not used</b>
<b>Type_Unit</b>	<b>System Type</b>	<b># Per/Sqd</b>	<b>not used</b>
<b>Type_Tactical</b>	<b>not used</b>	<b>Enabled</b>	<b>Unit type will appear as a firer or target</b>
<b>Unit_Code the model</b>	<b>Tracking number used by</b>	<b>potential</b>	
<b>Sensor</b>	<b>Sensor type available</b>		
<b>Mun_Name</b>	<b>Munition type available</b>		
<b>Mun_Code the model</b>	<b>Tracking number used by</b>		
<b>Mun_Storage type</b>	<b>Starting or basic load for munition</b>		
<b>Mun_Cargo</b>	<b>Munitions above the basic load</b>		
<b>Def_Type</b>	<b>Defensive munition type</b>		
<b>Def_Mun_Name</b>	<b>name of defensive munition</b>		
<b>Def_Mun_Code</b>	<b>Tracking number used by the model</b>		
<b>Num_Def_Muns</b>	<b>Starting or basic load for defense munitions</b>		
<b>Radios</b>	<b>Radio type available</b>		

# **Sample File Types (7 of 11)**

---

**The TRAC WSMR team creates various tables manually. These tables aid the analyst by allowing different formatting and data options to be included in the output. The following are sample files used to format output.**

**Manually generated database tables:**

- Category of Systems (CAS)**
- Category of Munitions (CAM)**
- Force Structure**
- Category of Systems by Force Structure (CAF)**

# Sample File Types (8 of 11)

---

**Category of Systems (CAS) table allows the analyst to combine system types into categories, e.g. Infantry and Armor.**

<u>Field</u>	<u>Description</u>
Type_Unit	System type of unit, populated from the Unit Summary File
Side	Side of system type
System_Type	Category name of system type
HVA	not used
HPT	not used
ID	Primary Key

**Category of Munitions (CAM) table defines the various munitions into categories of Line of Sight (LOS)/Non Line of Sight (NLOS) and Precision and Non-precision.**

<u>Field</u>	<u>Description</u>
ID	Primary Key
Mun_Name	Name of munition
NLOS	True/False value describing if the munition is LOS or NLOS
PREC	True/False value describing if the munition is precision guided or convention
Assigned	Shows if the munition has been assigned to a category

# Sample File Types (9 of 11)

---

**Force Structure table is populated by the unit list table. Extracts meaning from the units bumper number as to what brigade, battalion, company, platoon, and squad the unit belongs to.**

<u>Field</u>	<u>Description</u>
Side	Side of system type
UnitID	Bumper number of unit
Prefix	Classification of system type, e.g Unit TNF12345 - Prefix - TNF
Brigade	Brigade the unit belongs to, Unit TNF12345 - Brigade 1
Battalion	Battalion the unit belongs to, Unit TNF12345 - Battalion 2
Company	Company the unit belongs to, Unit TNF12345 - Company 3
Platoon	Platoon the unit belongs to, Unit TNF12345 - Platoon 4
Squad	Squad the unit belongs to, Unit TNF12345 - Squad 5

**Category of Systems by Force Structure (CAF) table allows the analyst to combine system types into categories, e.g. Infantry and Armor, and by brigade.**

<u>Field</u>	<u>Description</u>
ID	Primary Key
Side	Side of system type
Brigade	Brigade of system type
Type_Unit structure files	System type of unit, populated from the Unit Summary and force
System_Type	Category name of system type

# Sample File Types (10 of 11)

---

The **Tactical Area** table is a list of X and Y model coordinates defined as battlefield areas of interest (BAI). BAIs are defined by a lower left X and Y and an upper right X and Y creating a box.

<u>Field</u>	<u>Description</u>
LLX	X coordinate of the lower left position
LLY	Y coordinate of the lower left position
XSize	Distance in meters along the X axis the BAI extends from the LLX start point
YSize	Distance in meters along the Y axis the BAI extends from the LLY start point
Text	not used
NF	No Fire Zone.
BAI	Name of the current BAI

# **Sample File Types (11 of 11)**

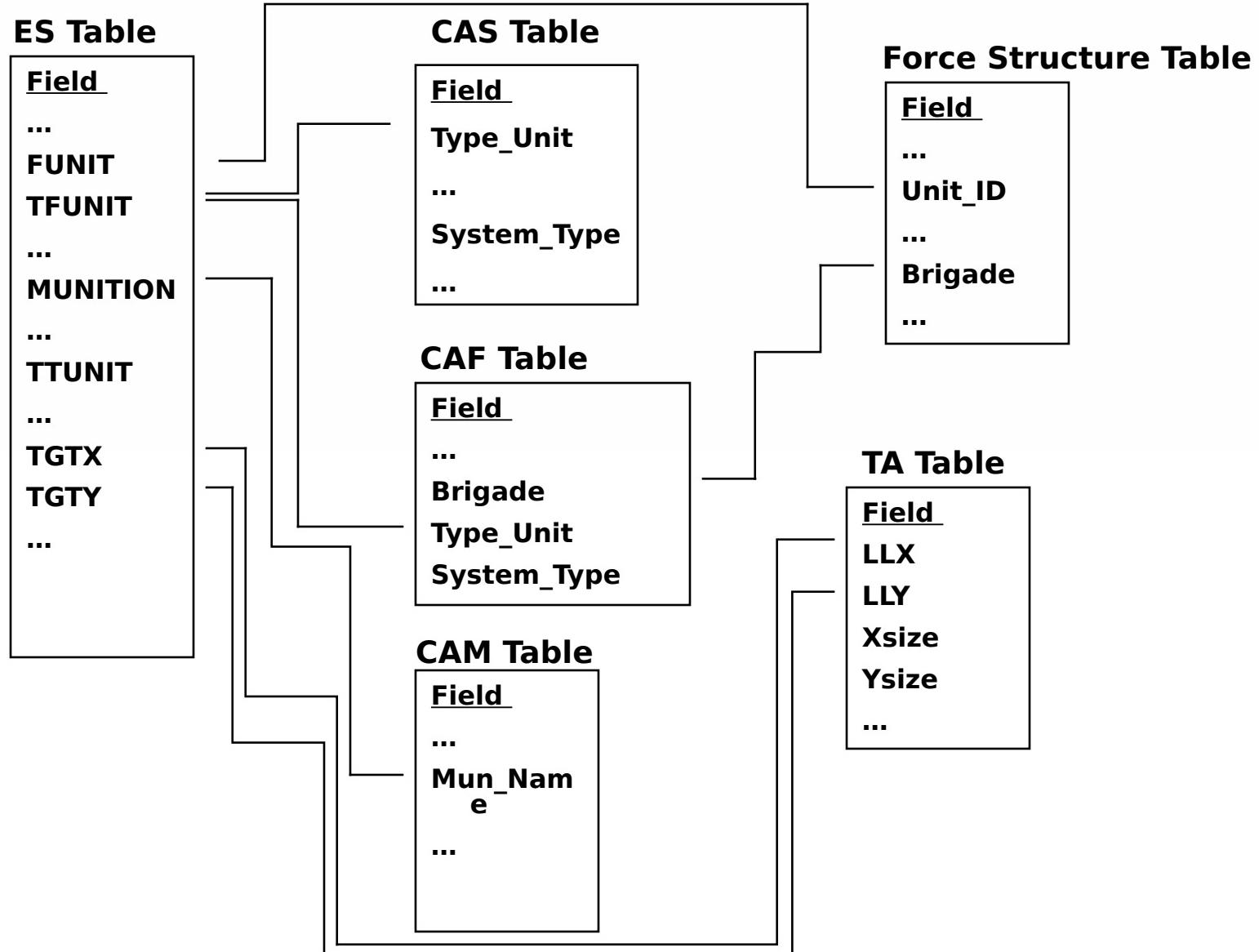
---

**Depending on the requirements of the analysis:**

- New files are created.**
- Old files are deleted.**
- File formats change to incorporate added resolution.**
  - An example of this is to add a column to show the ACQ and kill levels in the MT file.**

**!!!Any tool needs to have the flexibility to handle this!!!**

# File Types Relationships



... Denotes skipped fields. To view the full field headers refer to the previous slides.

# Sample MoEs (1 of 4)

---

**TRAC WSMR runs a multitude of queries to determine Measures of Effectiveness (MoE) over different tables. The following are an extremely small subset of potential MoEs:**

- **ES Table MoEs:**

- Threat shots at blue manned ground vehicles by aspect angle -30° interval.
- Blue infantry system kills of threat over time - 15 minute interval.
- Blue personnel kills by non line of sight (NLOS) system by range (100 meter interval).
- Unique acquisitions of threat artillery by blue FCS systems.
- Total acquisitions of threat high priority targets (HPT) by Blue Joint systems.
- Blue dismounted infantry losses by replication.
- Number of threat killed in BAI 03 between phase I and phase II.
- Number of threat shots along route 01 after phase II begins.
- All shots, hits, system kills, personnel kills, unique and total acquisitions of threat by blue systems.
- All shots, hits, system kills, personnel kills, unique and total acquisitions of blue by threat systems.

# Sample MoEs (2 of 4)

---

- **MT Table MoEs:**
  - **Average Location Error and Dispersion Error for Each Thread.**
  - **Number of Unique Targets Acquired.**
  - **Number of Unique Targets Acquired at Targetable Level.**
  - **Number of Unique Mission Threads Generated.**
  - **Number of Threat Units with Kill Assessments.**
  - **Quantity and Type of each message code, for each mission thread.**
  - **Time to achieve the desired outcome given a decision to employ an effect.**
  - **Time to achieve the desired outcome after the message for the first fire mission was received.**

# Sample MoEs (3 of 4)

---

## NOTE:

- **CAS, CAF, CAM, and TA tables are created by the user.** These tables need to be flexible and able to change.
- There can be multiple CAS, CAF, and TA tables. An example of this is a CAS table that breaks unit types out by “Infantry” and “Non Infantry” and another CAS table that breaks units out by “HPT”.
- **Force Structure table is populated by the UL table.** New units in the UL table must be added in the force structure table.
- The CAS, CAF, and CAM tables are populated by the US table. New systems and munitions in the US table must be added to these tables.
- **Phase I (360-1400), Phase II (1400-2475), Phase III (2475-2800)** all measured in minutes.
- **Unique Acquisition** is the first time it was acquired by the given constraints.

# Sample MoEs (4 of 4)

---

- **Average Location Error = Dist from Actual X,Y to Perceived X,Y.**
- **Average Dispersion Error = Dist from Perceived X,Y to Impact X,Y.**
- **Battle Damage Assessment (BDA) is a perceived effect against a target.**

# Requirements for Data Mining Tool

---

**TRAC WSMR requires any new tool to have the following functionality:**

- **Provide a system, which will be able to access and query over 500GB by 50 users simultaneously.**
- **Run extremely fast and complex queries in less than 3 minutes.**
- **Import text files into a database.**
- **Build relationships and queries (dynamically) between multiple files.**
- **Easily update, modify, and apply changes:**
  - **For model output file format changes and new tables.**
  - **Included software and analysis tools.**
  - **Added more resolution to the analysis as the models grow.**
- **Extract queries automatically with very little analyst interaction.**
- **Extract queries dynamically so analysts can specify what is important.**
- **Save queries and allow users to run a query over multiple data files.**
- **Format query results per user specifications with little analyst interaction.**
- **Present data in report (slide) format with little analyst interaction.**
- **Provide seamless communication between GUI, database engine, database files, and all software provided.**

# **Expected Deliverables**

---

**TRAC WSMR is planning to hold demos to review potential solutions. The following items should be adhered to:**

- **Show functionality with provided data files for Sept. 8<sup>th</sup> – 9<sup>th</sup> through a hands on demo with TRAC WSMR personnel.**
- **Show results for sample MoEs.**
- **Show various formatting capabilities for MoE output.**
- **Demonstrate capabilities for data mining and other analysis techniques using given data.**
- **Provide information on issues pertaining to:**
  - **cost in terms of hardware, software, installation, training, and overhead.**
  - **Purchasing man hours**
  - **Licensing**
  - **Software and analysis packages.**
    - **Statistical, mathematical, plotting capabilities, etc.**

# Contact Information

---

**Steven Ohman**

**(505) 678-8243**

**ohmansm@trac.wsmr.army.mil**

**Regina Berg**

**(505) 678-7089**

**bergrk@trac.wsmr.army.mil**

**Contact us for further MoE descriptions, CD for sample data, and any questions, comments, or concerns.**